Robust T-S Fuzzy Model Identification Approach Based on FCRM Algorithm and L1-Norm Loss Function

NAN ZHANG[1], **XIAOMING XUE**[1], **XIN XIA**[2], **WEI JIANG**[1], **CHU ZHANG**[2], **TIAN PENG**[2], **LIPING SHI**[1], **CHAOSHUN LI**[3], **(Member, IEEE), AND JIANZHONG ZHOU**[3]

[1]Jiangsu Key Laboratory of Advanced Manufacturing Technology, Huaiyin Institute of Technology, Huai'an 223003, China
[2]College of Automation, Huaiyin Institute of Technology, Huai'an 223003, China
[3]School of Hydropower and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China

Corresponding authors: Nan Zhang (zhangnanhust@163.com) and Xiaoming Xue (taxueli@hyit.edu.cn)

ABSTRACT The Takagi–Sugeno (T–S) fuzzy model identification is a very powerful tool for modelling of complicated nonlinear system. However, the traditional T-S fuzzy model typically uses the L2-norm loss function, which is very sensitive to outliers or noises. So an unreliable model may be obtained due to the presence of outliers or noises. In this paper, the outliers and noises robust T-S fuzzy model identification method based on the fuzzy c-regression model (FCRM) clustering and the L1-norm loss function is proposed. The hyper-plane-shaped clustering algorithm has been proved to be more effective than hyper-sphere-shaped clustering algorithm in T-S fuzzy model identification. Therefore the FCRM clustering algorithm is used in T-S fuzzy model identification for structure identification in the antecedent part. A mass of relevant researches have pointed out that the L1-norm loss function is more robust to outliers and noises than L2-norm loss function. In order to reduce the negative influence of outliers and noises, the L1-norm loss function is employed to enhance the robustness of T-S fuzzy model instead of the L2-norm loss function in the consequent part. Regression and classification applications have been used to demonstrate the validity of the proposed method. The experimental results show that the proposed method has significantly improved the modelling accuracy in dealing with data contaminated by outliers and noises compared with other algorithms.

INDEX TERMS T-S fuzzy model, fuzzy c-regression model, L1-norm loss function, outliers and noises robustness.

I. INTRODUCTION

With the rapid development of artificial intelligence techniques, the data-driven modelling method [1]–[7] is playing an important role for modelling of complicated nonlinear system in the age of big data, where the performance of these methods crucially rely on the quality of given training data. However, the training data will inevitably is contaminated by outliers or noises because of sampling errors, instrument errors and modelling errors, which will significantly reduce the accuracy and reliability of the established model. Therefore, it is very necessary to put forward the outliers and noises

robust data-driven modelling methods to handle the outliers and noises. Takagi–Sugeno (T–S) fuzzy model proposed by Takagi and Sugeno is a very effective data-driven modelling method [8]. It contains the following merits: (1) Easy-implementation [9], [10]; (2) High interpretability [11], [12]; (3) Universal approximation capability [13]–[15]. Therefore the T-S fuzzy model has been successfully and widely applied in the field of high precision modelling and many control problems [16]–[19].

Although the T-S fuzzy model has many advantages, the performance will be heavily deteriorated due to the presence of outliers or noises. The traditional T-S fuzzy model identification approach with L2-norm loss function is very sensitive to the outliers or noises because the L2-norm

The associate editor coordinating the review of this manuscript and approving it for publication was Ligang Wu.

loss function is prone to be badly affected by outliers and noises [20]. The unreliable T-S fuzzy model will be obtained in handling data with outliers or noises, and its performance can't be guaranteed. However, until now the outliers and noises robust T-S fuzzy model identification approach has not yet been widely researched. So robust T-S fuzzy model identification is worth to study.

However, few researches are focused on the robustness of the T-S fuzzy model with outliers and noises. But related work on other data-driven modelling methods can provide reference for building a robust T-S fuzzy model. Some outliers and noises robustness methods have been proposed for regression, classification, fault diagnosis [21], [22] and so on. Deng *et al.* [23] introduced the weighted least square scheme and the regularization term to improve the robustness of the extreme learning machine. Zhang *et al.* [24] used a weighted function to enhance the robustness of the least square support vector machine model to outliers. Although the weighted strategy can improve the robustness of algorithms to some extent, there also remain defects that limit its practical performance. The appropriate weight estimation relies heavily on previously established T-S fuzzy model and has an important influence on the final output. But we can't guarantee to obtain a good previous established T-S fuzzy model.

Most of related researches have proved that the L1-norm loss function is more robust to outliers and noises due to its sparsity [26]–[28]. For the L2-norm loss function, the corresponding loss increases in square with the increase of errors. When training data contains outliers or noises, regression errors in the outliers will dominate the increase or decrease of the entire loss function value, which will cause the trained model at this time to be biased towards outliers. But the loss of L1-norm function increase linearly with the error, which is slower than L2-norm loss function. Some studies have proved that the L1-norm is more robust than L2-norm and it has been widely applied in handling outliers or noises.

The fuzzy c-regression model (FCRM) clustering algorithm has been proposed for structure identification of T-S fuzzy model, which is a hyper-plane-shaped clustering algorithm [29]–[38]. The FCRM still has certain limitations when facing complex datasets with various shapes, sizes, and densities. So some clustering algorithms which are capable of clustering of arbitrary shape have been proposed [39]–[47]. Although these clustering algorithms have better performance than FCRM for data with various shapes, sizes, and densities, the FCRM described by several linear regression models maybe more suitable for the T-S fuzzy model identification because the final T-S fuzzy model is consisted of multiple linear regression models. Furthermore, this paper mainly studies the robustness of the T-S fuzzy model for data with outliers and noises, and the major innovation of this paper is in consequent parameters identification no structure identification. So in this paper, the mostly used FCRM clustering algorithm is adopted for T-S fuzzy model identification.

The hyper-plane-shaped clustering algorithm FCRM has been shown to be more effective than hyper-sphere-shaped

clustering algorithm in T-S fuzzy model identification. Thus, a robust T-S fuzzy model identification approach based on the FCRM clustering algorithm and L1-norm loss function is proposed. The FCRM is used to partition the input-output data space to obtain the optimal data structure of the T-S fuzzy model in the antecedent part. With the obtained antecedent parameters, the L1-norm loss function which has more strong outliers or noises robustness is introduced to estimate the output error instead of the traditional L2-norm loss function in the consequent part. Four regression applications and a classification application have been used to verify its effectiveness. The experimental results show that the proposed method has significantly improved the modelling accuracy in handling data with outliers and noises.

The rest of this paper is organized as follows. In Section II, the background of T-S fuzzy model and FCRM clustering algorithm are introduced. In Section III, the novel robust T-S fuzzy model identification approach is proposed. Section IV shows experimental results and analysis. Section V presents the conclusion and future works.

II. TAKAGI–SUGENO FUZZY MODEL IDENTIFICATION
A. T-S FUZZY MODEL
Assuming a multiple-input–single-output (MISO) system $P(x, y)$ needs to be identified, in which $x \in R^M$ is the system input and $y \in R$ is the system output. The T-S fuzzy model of this system is consisted of several IF-THEN fuzzy rules:

Rule R_i: IF x_1 is A_1^i and,..., and x_M is A_M^i THEN

$$y^i = \theta_0^i + \theta_1^i x_1 + \cdots + \theta_M^i x_M \quad (1)$$

where $R_i(i = 1, 2, 3, \ldots, c)$ is the fuzzy rule, c is the number of fuzzy rules. M is the dimension of the input vector, $x = [x_1, x_2, ..., x_M]^T$ is the input vector of the fuzzy model, which is composed of system input and output variables. "T" is the transposition operator of matrix. y^i is the i_{th} sub-model output and $\{\theta_j^i, j = 0, ..., M\}$ is the consequent parameter of i_{th} sub-model

The final output of T-S fuzzy model is comprised of those sub-models as a form of weighted mean defuzzification:

$$\hat{y} = \frac{\sum_{i=1}^{c} w^i y^i}{\sum_{i=1}^{c} w^i} \quad (2)$$

where the weight w^i denotes the overall membership grade of input x belonging to the i_{th} sub-model. It can be calculated as:

$$w^i = \prod_{j=1}^{M} \mu_{A_j^i}(x_j) \quad (3)$$

where $\mu_{A_j^i}(x_j)$ is fuzzy membership grade of x_j belonging to fuzzy set A_j^i.

There are triangular membership function, trapezoid membership function, Gauss membership function and so on can be chosen according to specific problems and experimental results. The Gauss membership function which is mostly used

in T-S fuzzy model is chosen in this paper:

$$\mu_{A_j^i}(x_j) = \exp(-\frac{1}{2} \cdot (\frac{x_j - v_j^i}{\sigma_j^i})^2) \quad (4)$$

where v_j^i, σ_j^i represent the center and width of the membership function respectively.

B. FUZZY C-REGRESSION MODEL CLUSTERING ALGORITHM

To obtain the center and width of the membership function, clustering algorithms are employed for fuzzy space partition. The FCRM has been used to partition the input–output data.

There are N input–output data pairs (x_k, y_k) $(k = 1, \ldots, N)$ and they will be divided into c subclasses using the FCRM, in which the ith subclass data samples are described by a linear regression model:

$$\begin{aligned} y_k^i &= f^i(x_k, \omega_i) = a_1^i x_{k1} + a_2^i x_{k2} + \cdots + a_M^i x_{kM} + b_0^i \\ &= [x_k^T \, 1] \cdot \omega_i \end{aligned} \quad (5)$$

It is a hyper-plane function. Where $x_k = [x_{k1}, \cdots x_{kM}]^T$ is the kth input vector, "·" is the multiplication operator of two vectors and $\omega_i = [a_1^i, \cdots, a_M^i, b_0^i]^T$ is the parameter of the ith linear regression model.

The distance between the kth sample to the ith linear regression model is defined as follows:

$$d_{ik}(\omega_i) = (y_k - f^i(x_k, \omega_i))^2 \quad (6)$$

Therefore the objective function of FCRM is defined:

$$\begin{cases} \min J(U, \omega) = \sum_{k=1}^{N} \sum_{i=1}^{c} (u_{ik})^m d_{ik}(\omega_i) \\ s.t. \sum_{i=1}^{c} u_{ik} = 1, k = 1, 2, \cdots, N \end{cases} \quad (7)$$

where $m \in (1, \infty)$ is the fuzzy weighted exponent and it is often set to be 2. $u_{ik} \in [0, 1]$ is the fuzzy membership degree of kth sample belonging to the ith cluster.

The fuzzy membership degree u_{ik} and the linear regression model parameter ω_i can be obtained by minimizing the objective function Eq. (7) [29]:

$$u_{ik} = \frac{1}{\sum_{r=1}^{c} [d_{ik}(\omega_i)/d_{rk}(\omega_i)]^{2/(m-1)}} \quad (8)$$

$$\omega_i = [X^T D_i X]^{-1} X^T D_i y \quad (9)$$

where $X = [x_1, x_2, x_3, \ldots, x_N]^T$, $D_i = diag\{u_{i1}, u_{i2}, u_{i3}, \ldots, u_{iN}\}$, $y = [y_1, y_2, y_3, \ldots, y_N]^T$ is the actual system output. By iterating Eq. (8) and Eq. (9), the fuzzy membership degree u_{ik} can be obtained.

C. IDENTIFICATION OF ANTECEDENT PARAMETERS

The main task of identification of antecedent parameters is to obtain reasonable center and width of membership function for each sub-model. Once the fuzzy membership matrix has been obtained, the antecedent parameters composed of the center v_{ij} and width σ_{ij} of fuzzy membership function can be calculated:

$$v_{ij} = \frac{\sum_{k=1}^{N} u_{ik} x_{kj}}{\sum_{k=1}^{N} u_{ik}} \quad (10)$$

$$\sigma_{ij} = \sqrt{\frac{2 \sum_{k=1}^{N} u_{ik}(x_{kj} - v_{ij})^2}{\sum_{k=1}^{N} u_{ik}}} \quad (11)$$

D. IDENTIFICATION OF CONSEQUENT PARAMETERS

The consequent parameters can be calculated from the following matrix equation:

$$y = A\theta \quad (12)$$

where $\theta = [\theta_0^1, \ldots, \theta_M^1, \ldots, \theta_0^c, \ldots, \theta_M^c]^T$ is the consequent parameters of T-S fuzzy model and A is the coefficient matrix, which can be calculated according to Eqs. (2)-(4):

$$A = \begin{bmatrix} \lambda_1^1 & \lambda_1^1 x_{11} & \cdots & \lambda_1^1 x_{1M} & , \cdots, & \lambda_c^1 & \lambda_c^1 x_{11} & \cdots & \lambda_c^1 x_{1M} \\ \cdots & \cdots & \cdots & \cdots & , \cdots, & \cdots & \cdots & \cdots & \cdots \\ \lambda_1^N & \lambda_1^N x_{N1} & \cdots & \lambda_1^N x_{NM} & , \cdots, & \lambda_c^N & \lambda_c^N x_{NM} & \cdots & \lambda_c^N x_{NM} \end{bmatrix} \quad (13)$$

where λ_i^k is defined as:

$$\lambda_i^k = \frac{w^i}{\sum_{i=1}^{c} w^i} \quad (14)$$

The least square method is used to solve the matrix equation in (12) to obtain the consequent parameters vector θ:

$$\theta = (A^T A)^{-1} A^T y \quad (15)$$

where the symbol "−1" is the inverse operator of matrix.

III. ROBUST T-S FUZZY MODEL IDENTIFICATION

In order to improve the robustness of T-S fuzzy model to outliers and noises, the robust T-S fuzzy model identification approach based on the FCRM algorithm and L1-norm loss function (RTS-L1) is proposed in this paper.

A. WEIGHTED T-S FUZZY MODEL

The weighted T-S fuzzy model (WTS) identification method is used to compare with RTS-L1 in this paper. It mainly includes three stages: (1) Train the traditional T-S fuzzy model and calculate the training errors of samples. Because of the presence of outliers or noises, the obtained T-S fuzzy model may be inaccurate; (2) Carry out a weighted scheme for T-S fuzzy model based on the training errors in the first stage. The basic principle of the weighted scheme is that the samples with high training errors are given small weights. On the contrary, samples with low training errors are given large weights, so that the influence of outliers or noises can be reduced. (3) Retrain the T-S fuzzy model based on the weighted samples and update the consequent parameters.

The objective function of consequent parameters identification of the traditional T-S fuzzy model is reported as

following:

$$\begin{cases} \min \|e\|_2^2 \\ subject\ to\ y - A\theta = e \end{cases} \quad (16)$$

where $\|x\|_p = \sqrt[p]{\sum_1^n |x_i^p|}$ is the L$_P$ norm. By solving the above loss function the optimal solution θ can be obtained. The matrix form of the objective function shown in Eq. (16) can be rewritten as:

$$J(\theta) = (y - A\theta)^T (y - A\theta) \quad (17)$$

For the unconstrained optimization problem, the optimal solution can be explicitly obtained by setting the gradient of the objective function J to zero:

$$\begin{aligned} \frac{\partial J(\theta)}{\partial \theta} &= \frac{\partial((y - A\theta)^T (y - A\theta))}{\partial \theta} \\ &= \frac{\partial (y^T y - y^T A\theta - \theta^T A^T y + \theta^T A^T A\theta)}{\partial \theta} \\ &= A^T A\theta - A^T y \end{aligned} \quad (18)$$

By setting $\frac{\partial J(\theta)}{\partial \theta} = 0$, the θ can be obtained shown in Eq. (15).

For the objective function shown in Eq. (16), the weights of all samples are assigned to be 1. However, the main idea of the WTS is to assign weights to the samples based on training errors. The weighted objective function of WTS is described as:

$$\begin{cases} min \|We\|_2^2 \\ subject\ to\ y - A\theta = e \end{cases} \quad (19)$$

where $W = diag\{\delta_1, \delta_2, ..., \delta_N\}$ is the weights of samples. Similar to Eq. (16), the objective function shown in Eq. (19) can be rewritten as:

$$J_w(\theta) = (y - A\theta)^T W^2 (y - A\theta) \quad (20)$$

The optimal solution can be explicitly obtained by setting the gradient of the objective function J_w to zero:

$$\begin{aligned} \frac{\partial J_w(\theta)}{\partial \theta} &= \frac{\partial (y - A\theta)^T W^2 (y - A\theta)}{\partial \theta} \\ &= \frac{\partial (y^T W^2 y - y^T W^2 A\theta - \theta^T A^T W^2 y + \theta^T A^T W^2 A\theta)}{\partial \theta} \\ &= A^T W^2 A\theta - A^T W^2 y \end{aligned} \quad (21)$$

Similar to Eq. (15), the optimal solution θ has the following form:

$$\theta = (A^T W^2 A)^{-1} A^T W^2 y \quad (22)$$

Now, the main task is to determine the weights of samples, the reasonable weights will lead to more reliable robust T-S fuzzy model. There are some methods to estimate the

weights based on the model training errors. Suykens et al. [48] proposed a classical weights estimation method:

$$\delta_i = \begin{cases} 1, & |e_i/\hat{s}| \le c_1 \\ \dfrac{c_2 - |e_i/\hat{s}|}{c_2 - c_1}, & c_1 < |e_i/\hat{s}| < c_2 \\ 10^{-4}, & otherwise \end{cases} \quad (23)$$

where $\hat{s}$ is the robust estimate of the standard variance of training errors, which has the form as follows:

$$\hat{s} = \frac{IQR}{2 \times 0.6745} \quad (24)$$

where IQR is the interquartile range of the samples errors. For input vector, IQR is the difference between the 75th and 25th percentiles of samples errors e. The values of c_1 and c_2 are typically set as 2.5 and 3.

B. ROBUST T-S FUZZY MODEL WITH FCRM AND L1-NORM LOSS FUNCTION

The majority researches have pointed out that the L1-norm loss function is more robust to outliers and noises than L2-norm loss function because of its sparsity. Therefore the L1-norm loss function is employed for identification of consequent parameters and the robust T-S fuzzy model based on FCRM and L1-norm loss function (RTS-L1) is proposed in this paper. The L1-norm loss function of the RTS- L1 has the form as follows:

$$\begin{cases} \min_{\theta} \|e\|_1 \\ subject\ to\ y - A\theta = e \end{cases} \quad (25)$$

Eq. (25) is a constrained convex optimization problem and it can be solved by the Augmented Lagrange Multiplier (ALM) method [49]. The corresponding augmented Lagrangian function is defined as [50]:

$$\begin{aligned} L_\mu (e, \theta, \lambda) \\ = \|e\|_1 + a^T (y - A\theta - e) + \mu/2\|y - A\theta - e\|_2^2 \end{aligned} \quad (26)$$

where g is a vector of Lagrange multiplier and b is a constant penalty factor. The ALM algorithm is used to obtain the optimal values of (e, θ, g) by iteratively updating the three parameters:

$$\begin{cases} \theta_{k+1} = (A^T A)A^T (y - e_k + g_k/b) \\ e_{k+1} = S_{1/\mu}[y - A\theta_{k+1} + g_k/b] \\ g_{k+1} = g_k + b(y - A\theta_{k+1} - e_{k+1}) \end{cases} \quad (27)$$

The function S_α is a shrinkage operator, which has the form as follows:

$$S_\alpha[x] = sign(x) \cdot \max\{|x| - \alpha, 0\} \quad (28)$$

Based on the above analysis, the detailed process of the proposed robust T-S fuzzy model identification approach with FCRM and L1-norm loss function is presented as follows:

Step 1: Data preprocessing. In order to make sure sample features are on a similar scale. The min-max Normalization
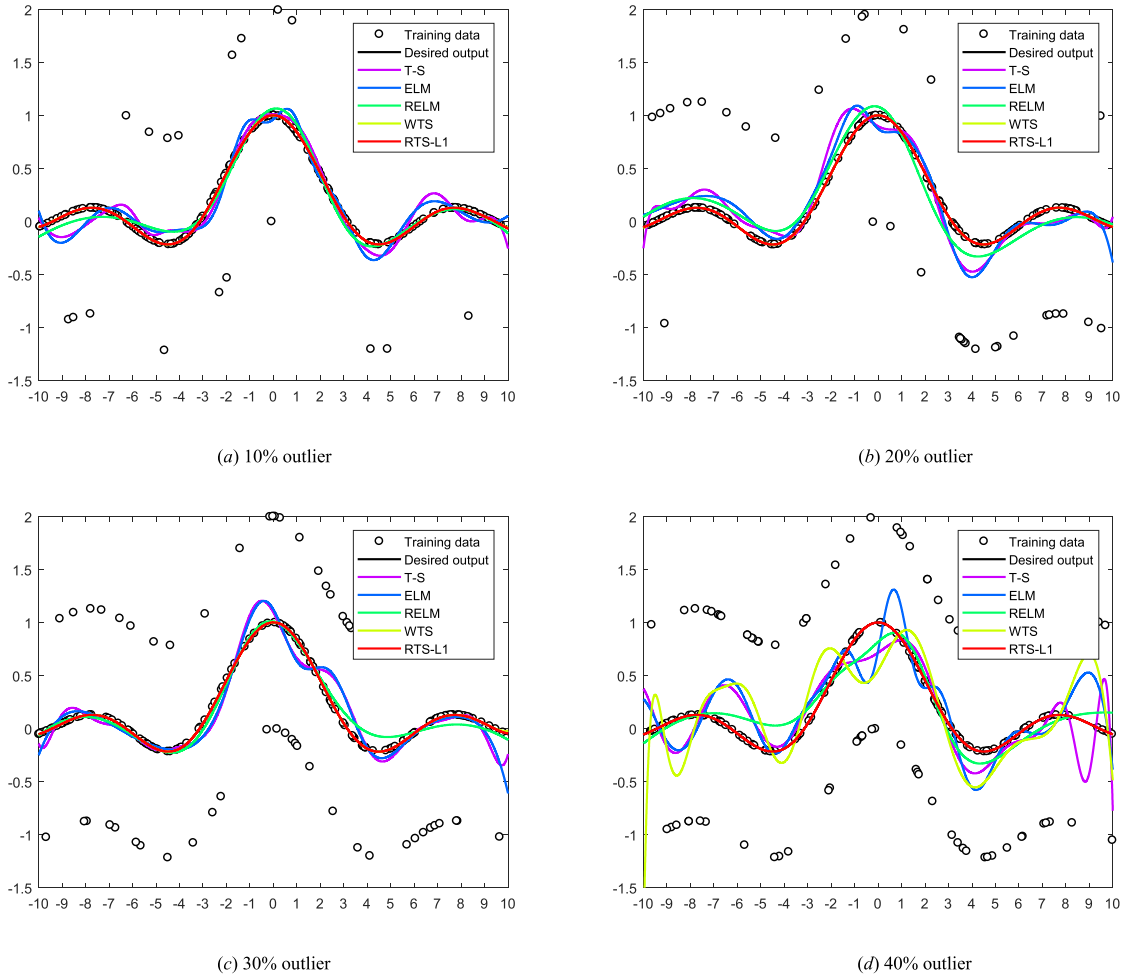
FIGURE 1. Predictive curves on training data contaminated by different proportion $\{-1, 1\}$ outlier.

is used to normalize the original data set to $[-1, 1]$. Divide it into training set and testing set and initialize the maximum number of iterations MaxIter $= 20$.

Step 2: Identification of antecedent parameters. Partition the input-output space of the T–S fuzzy model and calculate the fuzzy membership matrix based on the FCRM clustering algorithm. Identify the antecedent parameters according to Eqs. (10)-(11).

Step 3: Calculate the matrix $P_1 = (A^T A)^{-1} A^T$ and initialize parameters $e_1 = 0$, $g_1 = 0$;

Step 4: For $k = 1$ to MaxIter do:

Step 4.1: $\theta_{k+1} = (A^T A) A^T (y - e_k + g_k/b)$

Step 4.2: $e_{k+1} = S_{1/\mu}[y - A\theta_{k+1} + g_k/b]$

Step 4.3: $g_{k+1} = g_k + b(y - A\theta_{k+1} - e_{k+1})$

Step 4.4: $k = k + 1$

Step 4.5: end For.

Step 5: The consequent parameters vector θ can be obtained.

IV. EXPERIMENTS AND DISCUSSION

In this section, regression and binary classification applications are used to verify the performance of the proposed

RTS-L1 compared with the traditional T-S fuzzy model algorithm (T-S), extreme learning machine (ELM), regularized extreme learning machine (RELM) [51] and weighted T-S fuzzy model (WTS). The outliers and noises are added into training data to test the robustness of the proposed algorithm. It should be pointed out that outliers or noises only are added to training data not testing data. In this paper, the Root Mean Square Error (RMSE) is used as the performance index, which is defined as.

$$RMSE = \sqrt{\frac{\sum_{k=1}^{N} (y_k - \hat{y}_k)^2}{N}} \quad (29)$$

where y_k is the k_{th} original system output, and $\hat{y}_k$ is the k_{th} model output.

A. ROBUSTNESS EVALUATION VIA REGRESSION APPLICATION

In this part, the proposed RTS-L1 is compared with T-S, ELM, RELM and WTS by using four regression problems: SinC function, nonlinear differential equation, Box-Jekins system and Mackey–Glass chaotic time series. The experiments are carried out in the Matlab 9.2.0.538062 (R2017a)
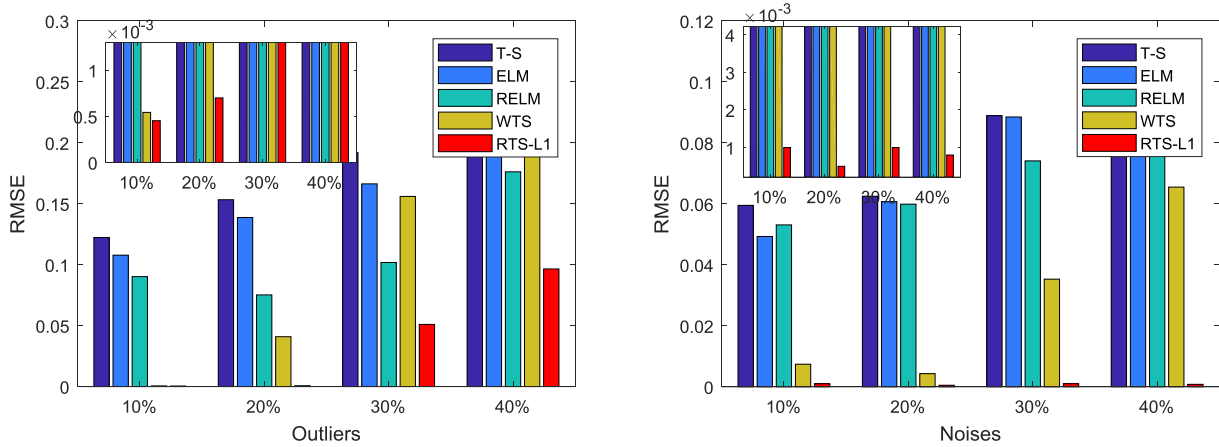
FIGURE 2. Comparison of testing RMSE of five algorithms for {−1, 1} outliers and [−1, 1] noises on SinC function.

TABLE 1. Training and testing RMSE of f ive algorithms for {−1, 1}outliers on SinC function.

Algorithms	10%		20%		30%		40%	
	Train	Test	Train	Test	Train	Test	Train	Test
T-S	0.2952	0.1223	0.4305	0.1533	0.5221	0.1920	0.5853	0.2615
ELM	0.2993	0.1078	0.4283	0.1387	0.5252	0.1663	0.5810	0.2642
RELM	0.3087	0.0902	0.4405	0.0752	0.5404	0.1018	0.6052	0.1762
WTS	0.3163	5.43e-4	0.4435	0.0409	0.5328	0.1560	0.5866	0.2587
RTS-L1	0.3162	**4.53e-4**	0.4473	**0.0007**	0.5440	**0.0510**	0.6268	**0.0965**

TABLE 2. Training and testing RMSE of five algorithms for [−1, 1] noises on SinC function.

Algorithms	10%		20%		30%		40%	
	Train	Test	Train	Test	Train	Test	Train	Test
T-S	0.1390	0.0595	0.2204	0.0625	0.3014	0.0889	0.3269	0.1054
ELM	0.1412	0.0493	0.2206	0.0607	0.3015	0.0885	0.327	0.0981
RELM	0.1495	0.0531	0.2253	0.0599	0.3091	0.0741	0.3343	0.0825
WTS	0.1473	0.0074	0.2268	0.0043	0.3116	0.0353	0.3322	0.0655
RTS-L1	0.1487	**0.0010**	0.227	**0.0005**	0.3130	**0.0010**	0.3407	**0.0008**

environment running in Intel (R) Core(TM) i5-3230M processor with the speed of 2.6 GHz. All experiments are repeated 30 times and average errors are recorded.

1) SINC FUNCTION
The regression problem of SinC function is often used to test the approximation ability of algorithm, which has the following mathematical expression [52]:

$$y = \begin{cases} \sin x/x, & x \neq 0 \\ 1, & x = 0 \end{cases} \qquad (30)$$

The 200 samples and 10001 samples are generated as training set and testing set respectively, in which x used for training and testing are generated from the uniform disturbance [−10, 10]. The input variable of the fuzzy model is $x(k)$ and the output variable is $y(k)$. The number of fuzzy rules is set to be 10. Firstly the outliers are added into the training set, which is randomly selected from the set {−1, 1} and

interval [−1, 1]. In order to test the robustness of the proposed algorithm with the increase of the number of outliers points, the 10%, 20%, 30% and 40% outliers or noises are taken into consideration.

Fig. 1 shows the results of regression on testing data using T-S, ELM, RELM, WTS and RTS-L1 on 10%, 20%, 30% and 40% {−1, 1} outliers level. The T-S and ELM are prone to be badly affected by outliers, especially for T-S. The RELM and WTS fit the testing data better than T-S and ELM on low outliers level (10%-30%), but the WTS arises fluctuations on 40% outliers level. The proposed RTS-L1 fits the testing data best than other four algorithms.

The experiment is repeated 30 times and the average training and testing RMSE are given in Table 1. For the purpose of comparing the performance of the five algorithms intuitively, the bar chart of testing RMSE are shown in Fig. 2.

From Table 1 and Fig. 2, the traditional T-S makes no distinction between outliers or noises and normal data, so it

TABLE 3. Training and testing RMSE of five algorithms for {−1, 1} outliers on NONLINEAR DIFFERENTIAL EQUATION.

Algorithms	10%		20%		30%		40%	
	Train	Test	Train	Test	Train	Test	Train	Test
T-S	0.3406	0.0475	0.4789	0.0981	0.5836	0.1166	0.6720	0.1716
ELM	0.3529	0.1030	0.4791	0.1154	0.5833	0.1071	0.6697	0.1829
RELM	0.3534	0.1036	0.4791	0.1187	0.5835	0.1090	0.6697	0.1765
WTS	0.3472	0.0344	0.4838	0.0608	0.5869	0.1481	0.6725	0.1548
RTS-L1	0.3475	**0.0320**	0.4884	**0.0549**	0.5987	**0.0788**	0.6781	**0.1284**

TABLE 4. Training and testing RMSE of five algorithms for [−1, 1] noises on NONLINEAR DIFFERENTIAL EQUATION.

Algorithms	10%		20%		30%		40%	
	Train	Test	Train	Test	Train	Test	Train	Test
T-S	0.7140	0.1244	0.8699	0.2420	1.1365	0.2895	1.2881	0.3764
ELM	0.7281	0.1796	0.8701	0.2393	1.1334	0.3093	1.2966	0.3999
RELM	0.7284	0.1798	0.8705	0.2393	1.1326	0.3077	1.2973	0.4104
WTS	0.7290	0.0511	0.8940	0.0864	1.1661	0.1307	1.3103	0.1507
RTS-L1	0.7322	**0.0488**	0.9040	**0.0693**	1.1851	**0.0695**	1.3293	**0.0680**
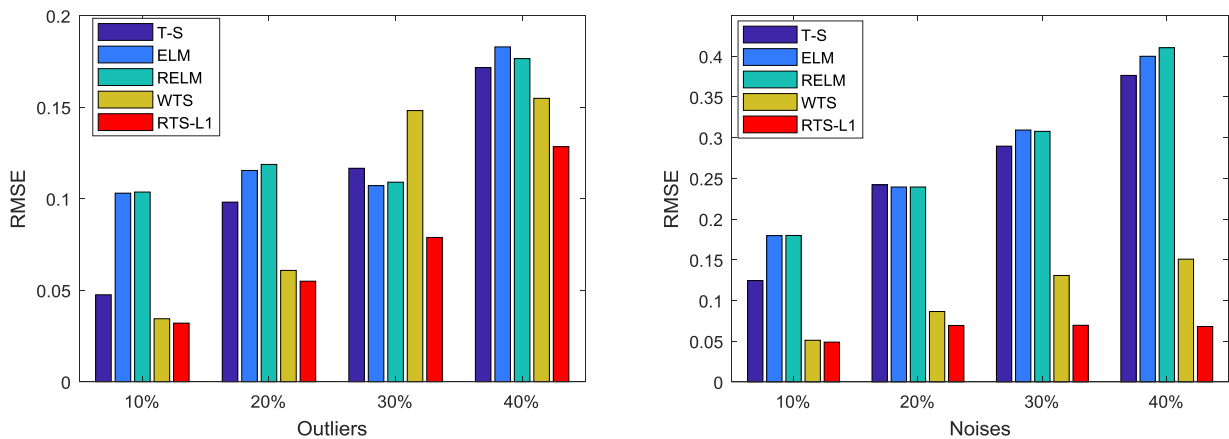
FIGURE 3. Comparison of testing RMSE of five algorithms for {−1, 1} outliers and [−1, 1] noises on nonlinear differential equation.

can obtain better training effect, but it fails for the testing process. The ELM and RELM have a better performance than T-S except for the ELM on 40% outliers level. The RELM obtains smaller testing RMSE than ELM due to its regularization term. Although the WTS obtains significant improvement than T-S, ELM and RELM, it doesn't obtain a better result than RELM on 30% and 40% {−1, 1} outliers. However, the proposed RTS-L1 obtains the best result on all outliers level.

Furthermore, the uniform distribution [−1, 1] noises are considered to add into training data for evaluating the noises robustness of algorithms. The experimental results with respect to [−1, 1] noises are shown in Table 2 and Fig. 2.

From Table 2 and Fig. 2, we can find that the T-S, ELM and RELM have obtained similar testing results, but the WTS and RTS-L1 gain significant improvement. However, the proposed RTS-L1 have improved at least one order of magnitude in testing RMSE than WTS.

From these experimental results, it is can be seen that the RTS-L1 have significantly improved the outliers and noises robustness of T-S fuzzy model.

2) NONLINEAR DIFFERENTIAL EQUATION
The nonlinear system is described by a second-order highly nonlinear difference equation [53]:

$$y(k) = \frac{y(k-1)y(k-2)(y(k-1)+2.5)}{1+y^2(k-1)+y^2(k-2)} + u(k) \quad (31)$$

The 500 data points are generated for training process and 500 data points are generated for testing process, where the uniform distribution in the interval [-2,2] for training data and a sinusoidal signal as $u(k) = \sin(2k/25)$ for testing data. The number of fuzzy rules is set to be four for T-S, WTS and RTS-L1 algorithms. The fuzzy model has three inputs $u(k)$, $y(k$-1), $y(k$-2) and a output $y(k)$. Tables 3 and 4 show the average training and testing RMSE of the five algorithms with

TABLE 5. Training and testing RMSE of five algorithms for {−1, 1} outliers on Box-Jekins system.

Algorithms	10%		20%		30%		40%	
	Train	Test	Train	Test	Train	Test	Train	Test
T-S	0.2977	0.1646	0.4561	0.2342	0.536	0.3239	0.6316	0.3228
ELM	0.3102	0.2056	0.4547	0.2298	0.5541	0.2868	0.6373	0.2947
RELM	0.3101	0.2034	0.4535	0.2259	0.5531	0.2899	0.6388	0.2891
WTS	0.3128	0.1492	0.4674	0.1863	0.5447	0.3473	0.6331	0.3124
RTS-L1	0.3229	**0.1068**	0.4677	**0.1519**	0.5582	**0.2046**	0.6527	**0.2668**

TABLE 6. Training and testing RMSE of five algorithms for [−1, 1] noises on Box-Jekins system.

Algorithms	10%		20%		30%		40%	
	Train	Test	Train	Test	Train	Test	Train	Test
T-S	0.189	0.1435	0.2585	0.2054	0.3253	0.262	0.3807	0.2368
ELM	0.187	0.1445	0.2595	0.1959	0.3282	0.2397	0.3816	0.2315
RELM	0.1882	0.1422	0.2591	0.2021	0.3265	0.2342	0.3795	0.2343
WTS	0.1993	0.1345	0.2653	0.1829	0.3334	0.258	0.3864	0.2543
RTS-L1	0.2036	**0.1097**	0.269	**0.1643**	0.3408	**0.1828**	0.3882	**0.2269**
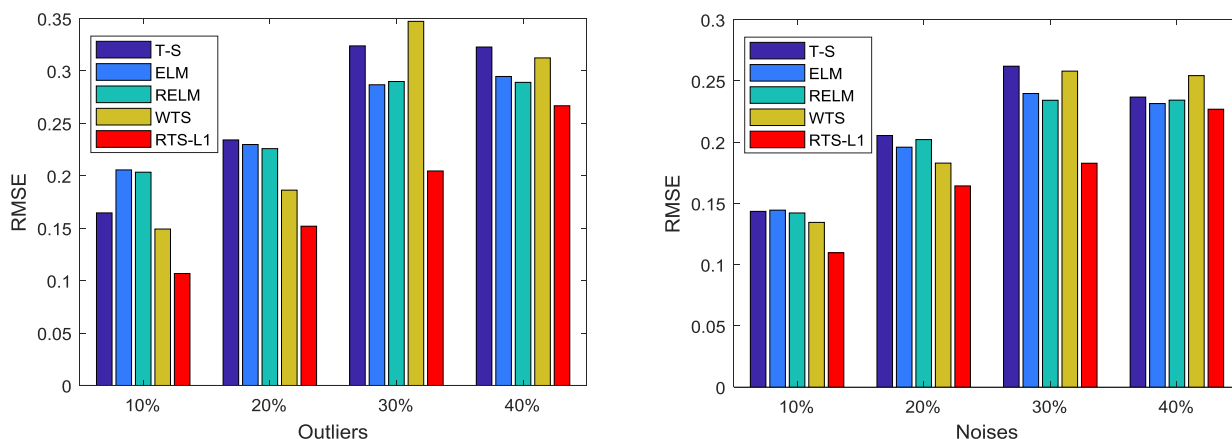
FIGURE 4. Comparison of testing RMSE of five algorithms for {−1, 1} outliers and [−1, 1] noises on Box-Jekins system.

TABLE 7. Training and testing RMSE of five algorithms for {−1, 1} outliers on Mackey–Glass chaotic time series.

Algorithms	10%		20%		30%		40%	
	Train	Test	Train	Test	Train	Test	Train	Test
T-S	0.3099	0.0657	0.4387	0.0828	0.5402	0.0882	0.6077	0.1757
ELM	0.31	0.0652	0.439	0.0873	0.538	0.1	0.6098	0.1628
RELM	0.3141	0.0476	0.4438	0.0673	0.543	0.0591	0.619	0.1056
WTS	0.3167	**0.0074**	0.4471	**0.0072**	0.5478	0.0103	0.6208	0.2856
RTS-L1	0.3165	0.0076	0.4469	0.0075	0.5478	**0.0101**	0.6316	**0.0061**

respect to {−1, 1} outliers and [−1, 1] noises on nonlinear difference equation data set, the corresponding bar chart of testing RMSE is compared in Fig. 3.

From Tables 3-4 and Fig. 3, the ELM and RELM have very similar testing RMSE for {−1, 1} outliers and [−1, 1] noises, and they fail to T-S in most cases. The WTS is better than T-S, ELM and RELM except for

30% {−1, 1} outliers. The proposed RTS-L1 has future improved the robustness and stability compared with the WTS. Take the result of [−1, 1] noises as an example, compared with the original T-S algorithm, the modelling accuracy of the proposed RTS-L1 has been promoted to be 0.0488, 0.0693, 0.0695 and 0.068 from 0.1244, 0.242, 0.2895 and 0.3764.

TABLE 8. Training and testing RMSE of five algorithms for [−5, 5] noises on Mackey–Glass chaotic time series.

Algorithms	10%		20%		30%		40%	
	Train	Test	Train	Test	Train	Test	Train	Test
T-S	0.7861	0.1503	1.2396	0.3081	1.4819	0.2623	1.849	0.2935
ELM	0.7885	0.1371	1.2391	0.3078	1.4829	0.2579	1.8432	0.3270
RELM	0.7913	0.1127	1.2498	0.2452	1.4918	0.1901	1.8551	0.2326
WTS	0.7999	0.0130	1.2715	0.0501	1.5025	0.0395	1.8693	0.0889
RTS-L1	0.8003	**0.0125**	1.2773	**0.0096**	1.5047	**0.0076**	1.8718	**0.0095**
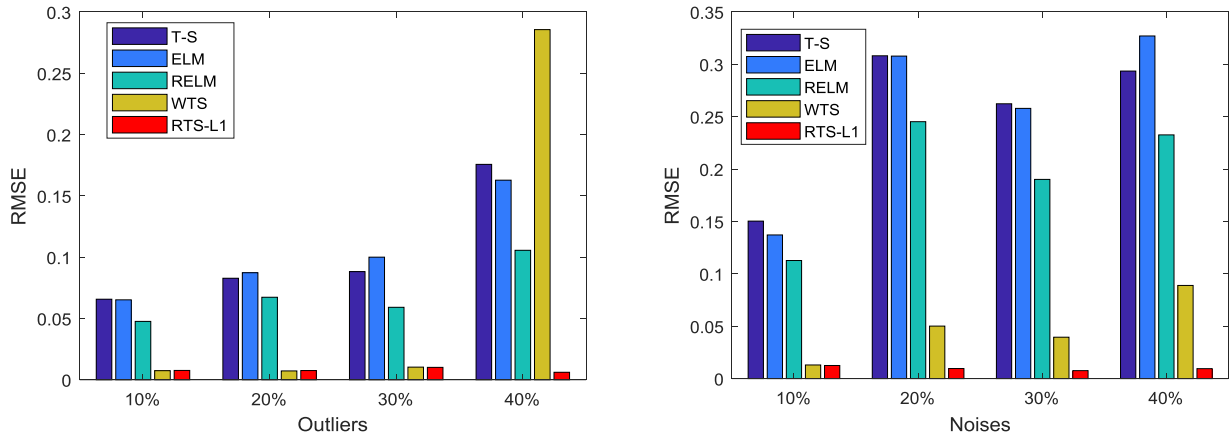
FIGURE 5. Comparison of testing RMSE of five algorithms for {−1, 1} outliers and [−1, 1] noises on Mackey–Glass chaotic time series.
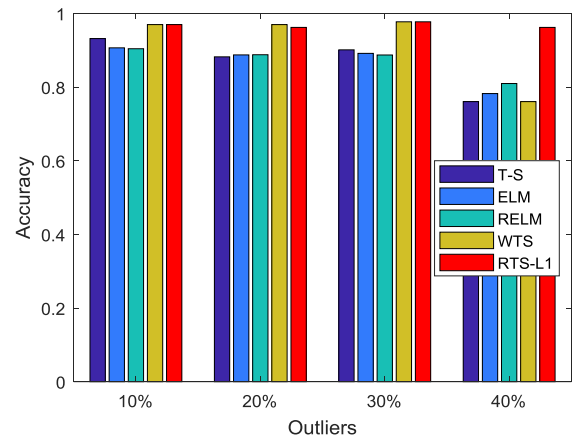
TABLE 9. Training and testing accuracy of five algorithms for {−1, 1} outliers on breast cancer data set.

Algorithms	10%		20%		30%		40%	
	Train	Test	Train	Test	Train	Test	Train	Test
T-S	0.8667	0.9316	0.8024	0.8821	0.7476	0.9011	0.4762	0.7605
ELM	0.8409	0.9065	0.7692	0.8873	0.7265	0.8915	0.5545	0.7823
RELM	0.8358	0.9041	0.7644	0.8878	0.7249	0.8871	0.5641	0.8096
WTS	0.9119	**0.9696**	0.8571	0.9696	0.8143	**0.9772**	0.4762	0.7605
RTS-L1	0.9071	**0.9696**	0.8429	**0.962**	0.8119	**0.9772**	0.7548	**0.962**

3) BOX-JEKINS SYSTEM

The Box-Jekins gas furnace system data set consisted of 296 data pairs describes the nonlinear relationship between the system input gas flow rate and the system output CO2 concentration [54]. The first 204 data pairs are considered as training set and the remaining data as testing set. The variable $y(k)$ is taken a fuzzy model output and the five variables $u(k)$, $u(k\text{-}1)$, $y(k\text{-}1)$, $y(k\text{-}2)$ and $y(k\text{-}3)$ are taken as fuzzy model input. The different proportion of {−1, 1} outliers and [−1, 1] noises are added to training set and the corresponding experimental results are shown in Tables 5-6 and Fig. 4.

Table 5 provides relevant comparison of training and testing RMSE of different algorithms on Box-Jekins system data with {−1, 1} outliers. The ELM and RELM which have the similar testing RMSE are better than T-S except for 10% outliers. The result of WTS is slightly better than T-S, ELM and RELM on 10% and 20% outliers, but worse on 30% and 40% outliers.

FIGURE 6. Comparison of testing RMSE of five algorithms on breast cancer data set on breast cancer data set.

Table 6 provides relevant comparison of training and testing RMSE of different algorithms on Box-Jekins system data
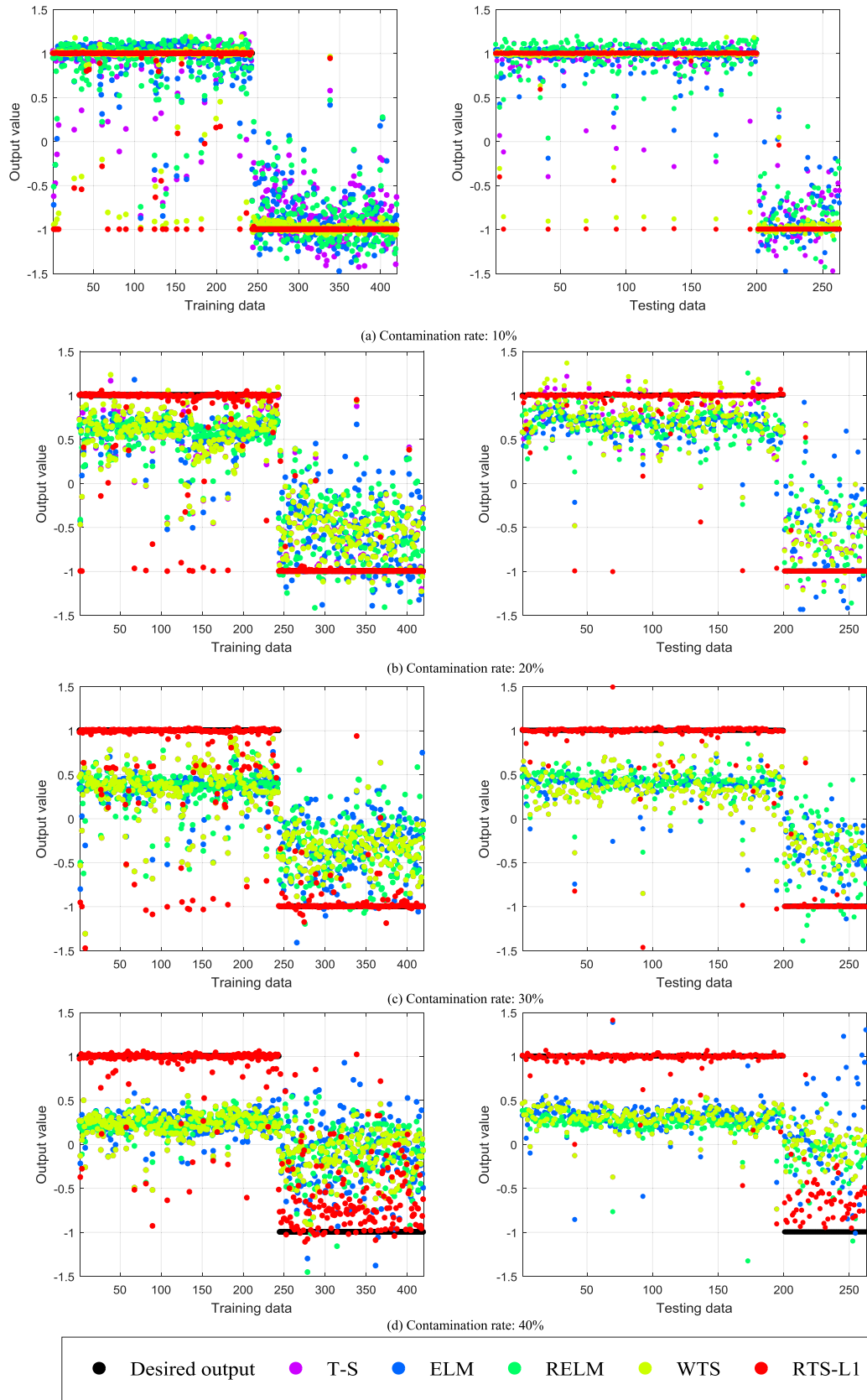
(a) Contamination rate: 10%

(b) Contamination rate: 20%

(c) Contamination rate: 30%

(d) Contamination rate: 40%

FIGURE 7. Training and testing outputs of five algorithms (contamination rate: 10%, 20%, 30% and 40%).

with [−1, 1] outliers. The T-S, ELM and RELM obtained a similar result. And the T-S also only obtained the better result on low outliers level. However, the RTS-L1 obtain the best result on 10%-40% outliers and noises level. It is superior to other methods listed in the table.

4) MACKEY–GLASS CHAOTIC TIME SERIES
As one of the famous time series predictive problems, the Mackey–Glass chaotic time series has been successfully and widely used to test the learning and generalization ability of different algorithms, which has the form as follows [55]:

$$\frac{dx(t)}{dt} = \frac{0.2x(t-17)}{1+x^{10}(t-17)} - 0.1x(t) \qquad (32)$$

There are 1000 input-output data pairs generated by the Eq. (32), where the first 500 data points are taken as the training set and the remaining 500 data points are taken as the testing set. The past four values $x(t\text{-}18)$, $x(t\text{-}12)$, $x(t\text{-}6)$ and $x(t)$ are used to predictive the future value $x(t\text{+}6)$ in the fuzzy model. All experiments are repeated 30 times and the average training RMSE and testing RMSE are shown in Tables 7-8 with respect to {−1, 1} outliers and [−5, 5] noises.

Table 7 shows that the WTS and RTS-L1 obtain smaller testing RMSE than T-S, ELM and RELM. The WTS and RTS-L1 have similar results on 10%, 20% and 30% outlier lever, but RTS-L1 significantly performs better than WTS when the outliers contamination rate is high (40%). From Table 8, the proposed RTS-L1 outperforms other algorithms on all noises level, especially in high noise level (20%, 30% and 40%). The bar chart of the testing RMSE of five algorithms is represented in Fig. 5.

B. ROBUSTNESS EVALUATION VIA BINARY CLASSIFICATION APPLICATION
The experimental results analysis of the proposed method for regression applications have been done above. In this part, a binary classification application is employed to verify the robustness of the proposed method. The Breast Cancer Wisconsin (Original) Data Set is used, which is consisted of 699 samples and each sample has 10 attributes and 2 possible classes: benign or malignant. In this data set, the benign is denoted by 1 and malignant is denoted by −1. Different from the regression applications, the outliers are added into training data set by changing the class of training samples. The different proportions of outliers are added into training data and the training and testing classification accuracy of five algorithms are shown in Table 9. Fig. 6 shows the corresponding classification accuracy of testing data for five algorithms.

From Table 9 and Fig. 6, the proposed RTS-L1 method has obtained higher classification accuracy for all outliers lever, but the WTS does not obtain expected results when the contamination rate is 40%. The training and testing outputs of T-S, ELM, RELM, WTS and RTS-L1 on 10%, 20%, 30% and 40% noises are shown in Fig. 7 (a)-(d). As can be seen from Fig. 7, the majority of outputs of RTS-L1 are

closely surround the desired output, but the outputs of other algorithms are more dispersed, especially for original T-S. Therefore the RTS-L1 shows more advantages than other algorithms.

The comparison experiments on regression problems and binary classification problem including of SinC function, nonlinear differential equation, Box-Jekins system, Mackey–Glass chaotic time series and Breast Cancer Wisconsin (Original) Data Set have been conducted. The T-S, ELM, RELM and WTS all use the L2-norm loss function as their objective function, but the L2-norm loss function is very sensitive to outliers or noises. The L1-norm loss function used in this paper is more robust to outliers or noises and the experiment results have demonstrated. The experimental results have shown that the proposed RTS-L1 is superior to T-S, ELM, RELM and WTS in most cases and show strong robustness for data with outliers and noises.

However, compared with T-S, ELM, RELM and WTS, the RTS-L1 has more points located in "output = −1" for some positive data in Fig. 7 and the WTS has the similar phenomenon. Although the most of outputs of RTS-L1 closely surround the desired outputs, the misclassification of RTS-L1 can prefer crisp values of {−1, 1} rather than fuzzy outputs. There still have certain limitations when the proposed method is applied in a binary classification application and the L1-norm loss function may be more suitable to deal with the regression problem with outliers or noises. Therefore the robust classification algorithm will be studied in our future work.

V. CONCLUSION AND FUTURE WORKS
In this paper, we have proposed the outliers and noises robust T-S fuzzy model identification method based on FCRM clustering algorithm and L1-norm loss function. The FCRM which is a hyper-plane-shaped clustering algorithm is used to obtain the more reasonable fuzzy space structure. The L1-norm loss function which has been proved to be more robust to outliers and noises than L2-norm loss function is employed in T-S fuzzy model identification instead of the L2-norm loss function. The augmented Lagrange multiplier iteration algorithm is used to solve the L1-norm loss function to guarantee its effectiveness and efficiency. Experimental results on several test problems including four regression applications and a classification application have shown that, compared with traditional T-S, ELM, RELM and WTS, the proposed RTS-L1 show more advantages on all datasets in dealing with data contaminated by outliers and noises.

Although the proposed RTS-L1 is more superior than other algorithms listed in this paper, it still has certain limitations that the misclassification of RTS-L1 can prefer crisp values of {−1, 1} rather than fuzzy outputs when it is applied in a binary classification application. Thus, our future work will focus on the following aspects:

(*a*) A robust classification algorithm will be studied in our future work.

(*b*) The intelligent optimization algorithm can be used to determine the optimal number of fuzzy rules [56]–[58].

REFERENCES

[1] W. Fu, K. Wang, C. Li, and J. Tan, "Multi-step short-term wind speed forecasting approach based on multi-scale dominant ingredient chaotic analysis, improved hybrid GWO-SCA optimization and ELM," *Energy Convers. Manage.*, vol. 187, pp. 356–377, May 2019.

[2] W. Fu, K. Wang, J. Tan, and K. Zhang, "A composite framework coupling multiple feature selection, compound prediction models and novel hybrid swarm optimizer-based synchronization optimization strategy for multi-step ahead short-term wind speed forecasting," *Energy Convers. Manage.*, vol. 205, Feb. 2020, Art. no. 112461.

[3] W. Fu, K. Shao, J. Tan, and K. Wang, "Fault diagnosis for rolling bearings based on composite multiscale fine-sorted dispersion entropy and SVM with hybrid mutation SCA-HHO algorithm optimization," *IEEE Access*, vol. 8, pp. 13086–13104, 2020.

[4] W. Fu, K. Wang, J. Tan, and K. Shao, "Vibration tendency prediction approach for hydropower generator fused with multi-scale dominant ingredient chaotic analysis adaptive mutation grey wolf optimizer and KELM," *Complexity*, vol. 2020, Jan. 2020, Art. no. 4516132.

[5] W. Fu, K. Wang, C. Zhang, and J. Tan, "A hybrid approach for measuring the vibrational trend of hydroelectric unit with enhanced multi-scale chaotic series analysis and optimized least squares support vector machine," *Trans. Inst. Meas. Control*, vol. 41, no. 15, pp. 4436–4449, Nov. 2019.

[6] X. Xue, C. Li, S. Cao, J. Sun, and L. Liu, "Fault diagnosis of rolling element bearings with a two-step scheme based on permutation entropy and random forests," *Entropy*, vol. 21, no. 1, p. 96, Jan. 2019.

[7] T. Peng, J. Zhou, C. Zhang, and Y. Zheng, "Multi-step ahead wind speed forecasting using a hybrid model based on two-stage decomposition technique and AdaBoost-extreme learning machine," *Energy Convers. Manage.*, vol. 153, pp. 589–602, Dec. 2017.

[8] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," in *Readings in Fuzzy Sets for Intelligent Systems*. Amsterdam, The Netherlands: Elsevier, 1993, pp. 387–403.

[9] C. Li, J. Zhou, J. Xiao, and H. Xiao, "Hydraulic turbine governing system identification using T–S fuzzy model optimized by chaotic gravitational search algorithm," *Eng. Appl. Artif. Intell.*, vol. 26, no. 9, pp. 2073–2082, 2013.

[10] L.-Q. Li, X.-L. Wang, W.-X. Xie, and Z.-X. Liu, "A novel recursive T–S fuzzy semantic modeling approach for discrete state-space systems," *Neurocomputing*, vol. 340, pp. 222–232, May 2019.

[11] Y. Ma and J. Zhao, "Cooperative output regulation for nonlinear multi-agent systems described by T–S fuzzy models under jointly connected switching topology," *Neurocomputing*, vol. 332, pp. 351–359, Mar. 2019.

[12] J. Chen, J. Li, and R. Duan, "T–S fuzzy model-based adaptive repetitive consensus control for second-order multi-agent systems with imprecise communication topology structure," *Neurocomputing*, vol. 331, pp. 176–188, Feb. 2019.

[13] S. Yan, J. Zhou, Y. Zheng, and C. Li, "An improved hybrid backtracking search algorithm based T–S fuzzy model and its implementation to hydroelectric generating units," *Neurocomputing*, vol. 275, pp. 2066–2079, Jan. 2018.

[14] Y. Zhao and H. Gao, "Fuzzy-model-based control of an overhead crane with input delay and actuator saturation," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 1, pp. 181–186, Feb. 2012.

[15] J. M. Adánez, B. M. Al-Hadithi, and A. Jiménez, "Multidimensional membership functions in T–S fuzzy models for modelling and identification of nonlinear multivariable systems using genetic algorithms," *Appl. Soft Comput.*, vol. 75, pp. 607–615, Feb. 2019.

[16] Y. Xu, Y. Zheng, Y. Du, W. Yang, X. Peng, and C. Li, "Adaptive condition predictive-fuzzy PID optimal control of start-up process for pumped storage unit at low head area," *Energy Convers. Manage.*, vol. 177, pp. 592–604, Dec. 2018.

[17] X. Wu, Y. Xu, J. Liu, C. Lv, J. Zhou, and Q. Zhang, "Characteristics analysis and fuzzy fractional-order PID parameter optimization for primary frequency modulation of a pumped storage unit based on a multi-objective gravitational search algorithm," *Energies*, vol. 13, no. 1, p. 137, Dec. 2019.

[18] Y. Xu, C. Li, Z. Wang, N. Zhang, and B. Peng, "Load frequency control of a novel renewable energy integrated micro-grid containing pumped hydropower energy storage," *IEEE Access*, vol. 6, pp. 29067–29077, 2018.

[19] Y. Zhao, J. Wang, F. Yan, and Y. Shen, "Adaptive sliding mode fault-tolerant control for type-2 fuzzy systems with distributed delays," *Inf. Sci.*, vol. 473, pp. 227–238, Jan. 2019.

[20] H. Wang, "Block principal component analysis with L1-norm for image analysis," *Pattern Recognit. Lett.*, vol. 33, no. 5, pp. 537–542, Apr. 2012.

[21] Y. Wu, B. Jiang, and N. Lu, "A descriptor system approach for estimation of incipient faults with application to high-speed railway traction devices," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 49, no. 10, pp. 2108–2118, Oct. 2019.

[22] Y. Wu, B. Jiang, and Y. Wang, "Incipient winding fault detection and diagnosis for squirrel-cage induction motors equipped on CRH trains," *ISA Trans.*, to be published.

[23] W. Deng, Q. Zheng, and L. Chen, "Regularized extreme learning machine," in *Proc. IEEE Symp. Comput. Intell. Data Mining*, Mar. 2009, pp. 389–395.

[24] C. Zhang, C. Li, T. Peng, X. Xia, X. Xue, W. Fu, and J. Zhou, "Modeling and synchronous optimization of pump turbine governing system using sparse robust least squares support vector machine and hybrid backtracking search algorithm," *Energies*, vol. 11, no. 11, p. 3108, Nov. 2018.

[25] C.-C. Chuang, S.-F. Su, J.-T. Jeng, and C.-C. Hsiao, "Robust support vector regression networks for function approximation with outliers," *IEEE Trans. Neural Netw.*, vol. 13, no. 6, pp. 1322–1330, Nov. 2002.

[26] P. Shi and L. Guoying, "On the rates of convergence of 'minimum l_1-norm' estimates in a partly linear model," *Commun. Statist.-Theory Methods*, vol. 23, no. 1, pp. 175–196, 1994.

[27] M. Daszykowski, K. Kaczmarek, Y. V. Heyden, and B. Walczak, "Robust statistics in data analysis—A review: Basic concepts," *Chemometric Intell. Lab. Syst.*, vol. 85, no. 2, pp. 203–219, 2007.

[28] K. Zhang and M. Luo, "Outlier-robust extreme learning machine for regression problems," *Neurocomputing*, vol. 151, pp. 1519–1527, Mar. 2015.

[29] R. J. Hathaway and J. C. Bezdek, "Switching regression models and fuzzy clustering," *IEEE Trans. Fuzzy Syst.*, vol. 1, no. 3, pp. 195–204, Aug. 1993.

[30] E. Kim, M. Park, S. Ji, and M. Park, "A new approach to fuzzy modeling," *IEEE Trans. Fuzzy Syst.*, vol. 5, no. 3, pp. 328–337, Aug. 1997.

[31] E. Kim, "A simply identified Sugeno-type fuzzy model via double clustering," *Inf. Sci.*, vol. 110, nos. 1–2, pp. 25–39, Sep. 1998.

[32] C. Li, J. Zhou, X. Xiang, Q. Li, and X. An, "T–S fuzzy model identification based on a novel fuzzy C-regression model clustering algorithm," *Eng. Appl. Artif. Intell.*, vol. 22, nos. 4–5, pp. 646–653, 2009.

[33] C. Li, J. Zhou, Q. Li, X. An, and X. Xiang, "A new T–S fuzzy-modeling approach to identify a boiler–turbine system," *Expert Syst. Appl.*, vol. 37, no. 3, pp. 2214–2221, Mar. 2010.

[34] C. Li, J. Zhou, L. Chang, Z. Huang, and Y. Zhang, "T–S fuzzy model identification based on a novel hyperplane-shaped membership function," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 5, pp. 1364–1370, Oct. 2017.

[35] C. Li, J. Zhou, B. Fu, P. Kou, and J. Xiao, "T–S fuzzy model identification with a gravitational search-based hyperplane clustering algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 2, pp. 305–317, Apr. 2012.

[36] C. Li, W. Zou, N. Zhang, and X. Lai, "An evolving T–S fuzzy model identification approach based on a special membership function and its application on pump-turbine governing system," *Eng. Appl. Artif. Intell.*, vol. 69, pp. 93–103, Mar. 2018.

[37] W. Zou, C. Li, and N. Zhang, "A T–S fuzzy model identification approach based on a modified inter type-2 FRCM algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 3, pp. 1104–1113, Jun. 2018.

[38] W. Zou, C. Li, and P. Chen, "An inter type-2 FCR algorithm based T–S fuzzy model for short-term wind power interval prediction," *IEEE Trans. Ind. Informat.*, vol. 15, no. 9, pp. 4934–4943, Sep. 2019.

[39] M. Chen, X. Wen, Z. Yang, M. Li, and M. Zhang, "MulSim: A novel similar-to-multiple-point clustering algorithm," *IEEE Access*, vol. 6, pp. 78225–78237, 2018.

[40] J. Huang, R. Xu, D. Cheng, S. Zhang, and K. Shang, "A novel hybrid clustering algorithm based on minimum spanning tree of natural core points," *IEEE Access*, vol. 7, pp. 43707–43720, 2019.

[41] L. Tang, W. H. Liu, S. Tang, Z. H. Niu, and J. C. Fan, "CDNASA: Clustering data with noise and arbitrary shape," *Int. J. Wireless Mobile Comput.*, vol. 11, no. 2, p. 100, 2016.

[42] B. K. Patra, S. Nandi, and P. Viswanath, "A distance based clustering method for arbitrary shaped clusters in large datasets," *Pattern Recognit.*, vol. 44, no. 12, pp. 2862–2870, Dec. 2011.

[43] F. Ros and S. Guillaume, "A hierarchical clustering algorithm and an improvement of the single linkage criterion to deal with noise," *Expert Syst. Appl.*, vol. 128, pp. 96–108, Aug. 2019.

[44] M.-C. Su and Y.-C. Liu, "A new approach to clustering data with arbitrary shapes," *Pattern Recognit.*, vol. 38, no. 11, pp. 1887–1901, Nov. 2005.

[45] L. Bai, J. Liang, and Y. Guo, "An ensemble clusterer of multiple fuzzy k-means clusterings to recognize arbitrarily shaped clusters," *IEEE Trans. Fuzzy Syst.*, to be published.

[46] H. Parvin, I. Jafari, and F. Rad, "A clustering ensemble learning method based on the ant colony clustering algorithm," *Int. J. Appl. Comput. Math.*, vol. 8, no. 3, p. 172, 2017.

[47] H. Parvin and B. Minaei-Bidgoli, "A clustering ensemble framework based on elite selection of weighted clusters," *Adv. Data Anal. Classification*, vol. 7, no. 2, pp. 181–208, Jun. 2013.

[48] J. A. K. Suykens, J. De Brabanter, L. Lukas, and J. Vandewalle, "Weighted least squares support vector machines: Robustness and sparse approximation," *Neurocomputing*, vol. 48, nos. 1–4, pp. 85–105, Oct. 2002.

[49] J. Yang and Y. Zhang, "Alternating direction algorithms for ℓ_1-problems in compressive sensing," *SIAM J. Sci. Comput.*, vol. 33, no. 1, pp. 250–278, Jan. 2011.

[50] L. Zhang, M. Yang, X. Feng, Y. Ma, and D. Zhang, "Collaborative representation based classification for face recognition," 2012, *arXiv:1204.2358*. [Online]. Available: http://arxiv.org/abs/1204.2358

[51] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 513–529, Apr. 2012.

[52] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, Dec. 2006.

[53] K. S. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks," *IEEE Trans. Neural Netw.*, vol. 1, no. 1, pp. 4–27, Mar. 1990.

[54] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*. Hoboken, NJ, USA: Wiley, 2015.

[55] J. Zhao and X. Yu, "Adaptive natural gradient learning algorithms for Mackey–Glass chaotic time prediction," *Neurocomputing*, vol. 157, pp. 41–45, Jun. 2015.

[56] X. Lai, C. Li, J. Zhou, Y. Zhang, Y. Li, "A multi-objective optimization strategy for the optimal control scheme of pumped hydropower systems under successive load rejections," *Appl. Energy*, vol. 261, Mar. 2020, Art. no. 114474.

[57] X. Lai, C. Li, J. Zhou, and N. Zhang, "Multi-objective optimization of the closure law of guide vanes for pumped storage units," *Renew. Energy*, vol. 139, pp. 302–312, Aug. 2019.

[58] Z. Zhao, J. Yang, W. Yang, J. Hu, and M. Chen, "A coordinated optimization framework for flexible operation of pumped storage hydropower system: Nonlinear modeling, strategy optimization and decision making," *Energy Convers. Manage.*, vol. 194, pp. 75–93, Aug. 2019.

XIN XIA received the Ph.D. degree in water conservancy and hydropower engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2015. He is currently a Lecturer with the Faculty of Automation, Huaiyin Institute of Technology, Huai'an, China. His research interests include modeling and fault diagnosis of hydro generator unit and system identification.

WEI JIANG received the B.S. degree in mechanical engineering and the Ph.D. degree in water resources and hydropower engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2014 and 2019, respectively. He is currently a Lecturer with the Faculty of Mechanical and Material Engineering, Huaiyin Institute of Technology. His research interests include fault diagnosis, trend prognosis, and intelligent management for mechanical system.
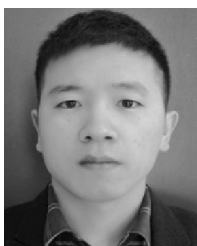
CHU ZHANG received the B.S. and Ph.D. degrees in water resources and hydropower engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2011 and 2018, respectively. He is currently a Lecturer with the College of Automation, Huaiyin Institute of Technology. His research interests include intelligent control and identification of hydraulic turbine unit and hardware circuit design in hydropower plants.

NAN ZHANG received the B.S. degree in electronic information engineering from Wuhan Textile University, Wuhan, China, in 2014, and the Ph.D. degree in water resources and hydropower engineering from the Huazhong University of Science and Technology, Wuhan, in 2019. He is currently a Lecturer with the Faculty of Mechanical and Material Engineering, Huaiyin Institute of Technology. His research interests include control theory, system identification, and modeling of power generation system.

TIAN PENG received the B.S. degree in information and calculation science from Chongqing University, Chongqing, China, in 2013, and the Ph.D. degree in system analysis and integration from the Huazhong University of Science and Technology, in 2018. She is currently a Lecturer with the College of Automation, Huaiyin Institute of Technology, Huaian, Jiangsu, China. Her main research interests include water resource management and optimization, energy forecasting, intelligence algorithms, and decision support system.

XIAOMING XUE received the B.S. degree in mechanical and electronic engineering from the Changchun University of Technology, Changchun, China, in 2010, and the M.S. and Ph.D. degrees in water resources and hydropower engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2012 and 2016, respectively.

He is currently a Lecturer with the Faculty of mechanical and material engineering, Huaiyin Institute of Technology. His research interests include vibration signal processing, intelligent fault diagnosis, and modeling of power generation system.
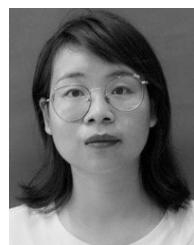
LIPING SHI is currently pursuing the M.Sc. degree with the Huaiyin Institute of Technology. Her research interests include control theory, system identification, and modeling of power generation system.
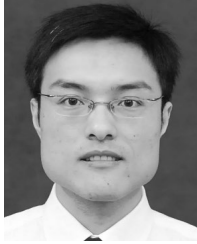
CHAOSHUN LI (Member, IEEE) received the
B.S. degree in thermal energy and power engi-
neering from Wuhan University, Wuhan, China,
in 2005, and the Ph.D. degree in water conservancy
and hydropower engineering from HUST, in 2010.
He is currently an Associate Professor with the
School of Hydropower and Information Engineer-
ing, HUST. His research interests include fuzzy
application system, fuzzy modeling, and system
identification.

JIANZHONG ZHOU was born in Wuhan, China,
in December 1959. He received the B.S. degree
in automatic control from the Nanjing University
of Aeronautics and Astronautics, Nanjing, China,
in 1982. He is currently a Professor with the
School of Hydropower and Information Engineer-
ing, HUST. His research interest includes mod-
eling, control, and operation theory in hydraulic
power plants.

● ● ●